

BEST AVAILABLE COPY
PATENT ABSTRACTS OF JAPAN

(11)Publication number : 63-298523
(43)Date of publication of application : 06.12.1988

(51)Int.Cl. G06F 7/28
G06F 15/66
G06F 15/70

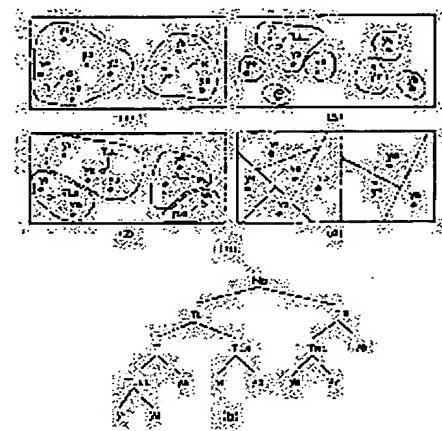
(21)Application number : 62-131362 (71)Applicant : RICOH CO LTD
(22)Date of filing : 29.05.1987 (72)Inventor : OGAWA OSAMU

(54) SYSTEM FOR FORMING BINARY TREE CONSTRUCTION DICTIONARY BY CLUSTERING METHOD

(57)Abstract:

PURPOSE: To improve a matching accuracy by distributing a reference vector strong in a correlation to the same cluster.

CONSTITUTION: When a state for dividing all the reference vectors $y_1, y_2 \dots y_n$ into two clusters TL, TR consisting of the reference vectors mutually strong in the correlation $y_{L1}, y_{L2} \dots y_{Lj}$ and $y_{R1}, y_{R2} \dots y_{Rk}$ is typically indicated, only the correlation is noticed without considering the number of the reference vectors included in the clusters TL, TR to execute the clustering. Thereby, an input vector and the reference vector constituted as a hierarchical binary tree are defined to be a hierarchical dictionary for matching in the respective nodes and the dictionary removing the generating of a matching distortion and having high matching accuracy.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

BEST AVAILABLE COPY

⑨ 日本国特許庁 (JP)

⑩ 特許出願公開

⑪ 公開特許公報 (A)

昭63-298523

⑫ Int. Cl.

G 06 F 7/28
15/66 15/70

識別記号

330
465

厅内整理番号

X-7313-5B
8419-5B
7368-5B

⑬ 公開 昭和63年(1988)12月6日

審査請求 未請求 発明の数 1 (全5頁)

⑭ 発明の名称 クラスタリング法による2進木構造辞書の作成方式

⑮ 特願 昭62-131362

⑯ 出願 昭62(1987)5月29日

⑰ 発明者 小川 修 東京都大田区中馬込1丁目3番6号 株式会社リコー内

⑱ 出願人 株式会社リコー 東京都大田区中馬込1丁目3番6号

⑲ 代理人 井理士 滝野 秀雄 外1名

明 細 書

1. 発明の名称

クラスタリング法による2進木構造辞書の作成方式

2. 特許請求の範囲

(1) 多数のK次元ベクトルデータを2進木構造により階層的に分類した辞書を作成するために、上記多数のK次元ベクトルデータの内で分類の対象となるデータをこれらデータ相互間の相関の大小のみに基づいて2分割して2進木構造辞書を作成することを特徴とするクラスタリング法による2進木構造辞書の作成方式。

(2) 分割されたデータ群を、このデータ群内に含まれるデータ相互間の相関の大小に基づいて2分割し、さらにこれを繰返して2進木構造辞書を作成することを特徴とする特許請求の範囲第1項記載のクラスタリング法による2進木構造辞書の作成方式。

3. 発明の詳細な説明

【技術分野】

入力した多次元ベクトルと基準ベクトルとを比較するマッチング処理によって、入力した多次元ベクトルを分類し、あるいは符号化、復号化を行うために入力ベクトルに最も近似した基準ベクトルを選択することが行われる。

上記マッチング処理の1つの方法として、基準ベクトルを各ノードにおいて2進木として順次階層的に構成し、これらノードにおいて入力ベクトルと基準ベクトルとを逐次比較することによって2進木の一方の入力ベクトルに近い当該ノードにおける基準ベクトルへの分岐を選択し、これを逐次のノードにおいて繰返すことによって入力ベクトルに最も近似した基準ベクトルを選択するようにした判定法に関する。

本発明は、上記のようなマッチング処理を行うために、階層的な各ノードにおいて2進木に分岐させるためのクラスタリング法による2進木構造辞書の作成方式に関する。

〔従来技術〕

第2図は $2^3 = 8$ 箇の基準ベクトル $\alpha_1, \alpha_2, \dots, \alpha_8$ からなる2進木の例を示すもので、第1層のノード $N_{1,1}, N_{1,2}$ におけるそれぞれの基準ベクトル $\alpha_{1,1}, \alpha_{1,2}$ は、これらノードに継続する下位ノード、すなわち第2層のノード $N_{2,1,1}, N_{2,1,2}$ における2つの基準ベクトル $\alpha_{2,1,1}, \alpha_{2,1,2}$ および第2層のノード $N_{2,2,1}, N_{2,2,2}$ における2つの基準ベクトル $\alpha_{2,2,1}, \alpha_{2,2,2}$ をそれぞれ代表するものであり、これら基準ベクトル $\alpha_{1,1}, \alpha_{1,2}$ の値として、後にさらに詳細に説明するように、上記の継続する下位ノードの2つの基準ベクトルの平均値を用いることが一般に行われている。

同図によってより具体的にいえば、ノード $N_{1,1}$ の基準ベクトル $\alpha_{1,1}$ としてはこのノード $N_{1,1}$ に継続する下位ノード $N_{2,1,1}, N_{2,1,2}$ のベクトル $\alpha_{2,1,1}$ と $\alpha_{2,1,2}$ の平均値を、また、ノード $N_{1,2}$ の基準ベクトル $\alpha_{1,2}$ としてはこのノードに継続する下位ノード $N_{2,2,1}, N_{2,2,2}$ のベクトル $\alpha_{2,2,1}$ と $\alpha_{2,2,2}$ の平均値を用いるものであり、第2図と第3層、第3

層と第4層というような、より下位の層についても同様な基準ベクトルが用いられる。

第3図はこのような2進木によるマッチングを行うための多數のベクトルの分類による辞書の従来の作成方法を概念的に示したもので、多數のK次元ベクトルの基準ベクトル、すなわち辞書データが存在する空間Sを実線の矩形で示してあり、これら基準ベクトルが各空間に同数ずつ配分されるようにこの例では上記8箇の空間に分割され、これら各空間ごとに当該空間に属するベクトルの例えれば平均値によってこの空間を代表する基準ベクトル $\alpha_1, \alpha_2, \dots, \alpha_8$ が生成される。

そして、これらの基準ベクトル $\alpha_1, \alpha_2, \dots, \alpha_8$ の内比較的近い基準ベクトル例えれば α_1, α_2 を1組として基準ベクトル $T_{1,1}$ を作成し、同様にして上記基準ベクトル α_3, α_4 から作成した基準ベクトル $T_{1,2}$ との平均値を次の上位のノード $N_{1,1}$ についての基準ベクトル $\alpha_{1,1}$ とする。

さらに、上記同様に基準ベクトル $\alpha_5, \alpha_6, \alpha_7, \alpha_8$ から作成された基準ベクトル $\alpha_{1,2}$ と上

記基準ベクトル $\alpha_{1,1}$ によってより上位のノード N_1 を生成させる。

このようにして作成される従来の辞書においては、分類された各群に属する基準ベクトルが同数になるように分割していたため、検索上の便宜はあるにしても、相互の相關の高い基準ベクトルが他の群内に含まれるようになり、探索の精度が低くなる問題点があった。

第5図(回)はこの従来技術における問題点を説明するためのもので、入力ベクトル又は $\#$ 側クラスタ内の基準ベクトル α に最も近いにもかかわらず、この $\#$ 側クラスタを代表する代表基準ベクトル T_1 の方が $\#$ 側クラスタを代表する代表基準ベクトル T_2 より入力ベクトル又に近いために、上記最も近い基準ベクトル α はマッチングの対象から外されてしまってこの基準ベクトル α を選択することが不可能になり、マッチング歪を生ずる原因となっている。

〔目的〕

本発明は、入力ベクトルと階層的な2進木として構成された基準ベクトルとを各ノードにおいてマッチングを行うための階層的な辞書として、上記のようなマッチング歪の発生を除いた、マッチング精度の高い辞書を得ることを目的とするものである。

〔構成〕

本発明のクラスタリング法を用いた2進木構造辞書の作成方式を、1実施例に基づいて以下説明する。

第1図は本発明によるクラスタリングによって、すべての基準ベクトル $\alpha_1, \alpha_2, \dots, \alpha_8$ を互いに相關の強い基準ベクトル $\alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{1,8}$ および $\alpha_{2,1}, \alpha_{2,2}, \dots, \alpha_{2,8}$ からなる2つのクラスタ K_1, K_2 に分割する状態を模式的に示すものであり、このときそれぞれのクラスタ K_1, K_2 に含まれる基準ベクトルの数は考慮しないで相關度のみに着眼してクラスタリングを行うことを特徴と

する。

同図(a)は8つの基準ベクトルデータ $\alpha_1, \alpha_2, \dots, \alpha_8$ を本発明によって2進木構造の辞書として作成する例を示すもので、その(1)図は第1階層のノードにおける上記基準ベクトルデータ $\alpha_1, \alpha_2, \dots, \alpha_8$ の分割を示しており、互いに相関の強い5つの基準ベクトルデータ $\alpha_1, \alpha_2, \dots, \alpha_5$ が左方のクラスタ T_L に、また互いに相関の強い3つの基準ベクトルデータ $\alpha_6, \alpha_7, \alpha_8$ が右方のクラスタ T_R に配分されており、これらクラスタ T_L, T_R の代表基準ベクトル α_L, α_R としては、例えばK-平均値法によってそれぞれのクラスタに属するベクトルの平均値を用いることができる。

第1図(a)の(2)図は次の第2階層での基準ベクトルデータ $\alpha_1, \alpha_2, \dots, \alpha_8$ の分割を示しており、上記左方のクラスタ T_L は基準ベクトル $\alpha_1, \alpha_2, \alpha_3$ からなる左方のクラスタ T_{LL} と基準ベクトル $\alpha_4, \alpha_5, \alpha_6$ からなる右方のクラスタ T_{LR} とに分割されており、このクラスタ T_{LL} の代表基準ベクトル α_{LL} は例えば上記同様に基準ベクトル $\alpha_1, \alpha_2, \alpha_3$

の平均値であり、またクラスタ T_{LR} の代表基準ベクトル α_{LR} はこのクラスタに含まれる基準ベクトル $\alpha_4, \alpha_5, \alpha_6$ の平均値となる。

また、上記右方のクラスタ T_R は基準ベクトル α_7, α_8 からなる左方のクラスタ T_{RL} と基準ベクトル α_8 からなる右方のクラスタ T_{RR} とに分割されており、このクラスタ T_{RL} の代表基準ベクトル α_{RL} は例えば上記同様に基準ベクトル α_7, α_8 の平均値であり、またクラスタ T_{RR} の代表基準ベクトル α_{RR} はこのクラスタに含まれる基準ベクトルが α_8 のみであることからこの基準ベクトル α_8 となる。

さらに、第1図(a)の(3)図は次の第3階層での基準ベクトルの分割を示しており、上記クラスタ T_{LL}, T_{LR} および T_{RR} がそれぞれ分割されて基準ベクトル $\alpha_1, \alpha_2, \alpha_3$ からなるクラスタのみが複数の基準ベクトルを含むクラスタとなっており、その他のクラスタはそれぞれ基準ベクトル $\alpha_4, \alpha_5, \alpha_6$ ごとに独立したクラスタとなる。

次の階層として、上記の基準ベクトル $\alpha_1, \alpha_2, \alpha_3$

からなるクラスタを分割すると、第1図(a)の(4)図に示したように、各基準ベクトルごとに分割されて、入力ベクトル又に最も近い基準ベクトルをこれら基準ベクトル $\alpha_1, \alpha_2, \alpha_3$ 中から選択することができる。

第1図(b)はこのようにして得られた2進木辞書の構造を示すもので、第2図に示した従来の2進木辞書の構造と対比すれば明らかなように、本発明では各クラスタに含まれる基準ベクトルの数を考慮することなく相関の強さのみによってクラスタを構成したため、他の基準ベクトルとの相関の低い基準ベクトル、上例では $\alpha_4, \alpha_5, \alpha_6$ は比較的高い階層においてマッチングが行われ、逆に他の基準ベクトルとの相関の強い基準ベクトル、上例では $\alpha_1, \alpha_2, \alpha_3$ は比較的深い階層においてマッチングが行われることになる。

第4図は本発明に基づき、コンピュータ処理によってN個の基準ベクトルを上記のような辞書に構成するためのフローチャートを示すもので、このN個の基準ベクトルにラベルを付して分類を行

うものであり、このラベルは基準ベクトル間の相関に基づき、第1階層においては1あるいは2のラベルのいずれかをこれら基準ベクトルのそれぞれに付し、第2階層では上記1のラベルが付された基準ベクトルを3, 4のラベルのいずれかに、また2のラベルが付された基準ベクトルを5, 6のラベルのいずれかに分類して付替える。

さらに、上記3～6のラベルは次の第3階層において7～14のラベルに上記同様に付替えることによって基準ベクトルをそれぞれ8つのクラスタのいずれかに配分し、これを繰返してすべての基準ベクトルを2進木辞書の末端に割当てるものである。

なお、これらラベルは上記処理の終了後は必要ないので、削除することができる。

第5図(b)は本発明によるマッチング歪の減少を説明するためのもので、従来例を示す同図(a)と比較すれば明らかなように、本発明によって相関の強い基準ベクトルは同一のクラスタに配分されるため、入力ベクトル又に最も近い基準ベクトル $\alpha_1, \alpha_2, \alpha_3$

は従来例とは異なって、の個クラスタに属しており、したがってこの基準ベクトルア。がマッチングの対象から外されることがないのでマッチングの精度を向上することができる。

[効 果]

本発明によれば、相間の強い基準ベクトルは同一のクラスタに配分されるため、入力ベクトルに最も近い基準ベクトルがマッチングの対象から外されることがないのでマッチングの精度を向上することができる。

特許出願人 株式会社 リコ一
代理人 橋野秀雄
同 中内康雄



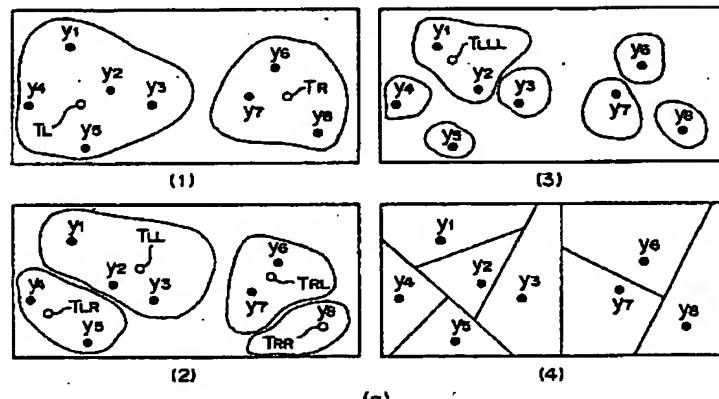
4. 図面の簡単な説明

第1図は本発明によるクラスタリングによって多数の基準ベクトルをクラスタに分割する状態を模式的に示す図、

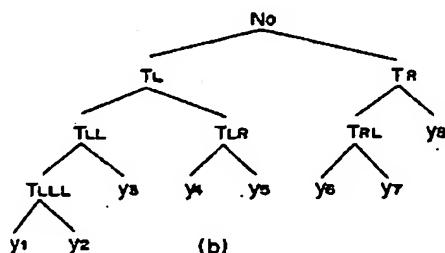
第2図は2進木によるマッチングの説明図、

第3図はこのマッチングを行うための2進木を生成する従来の方法を説明する図、

第4図は本発明による2進木構造辞書の作成方式

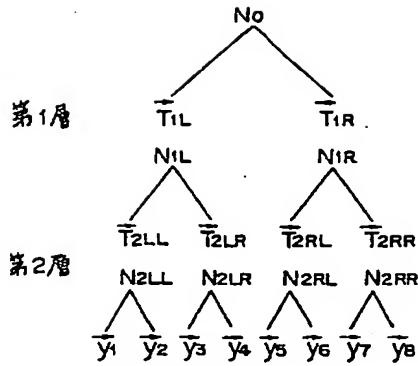


(a)

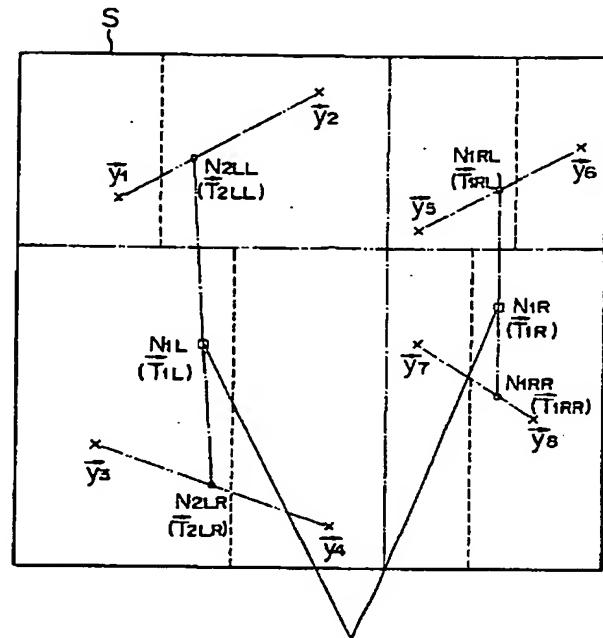


(b)

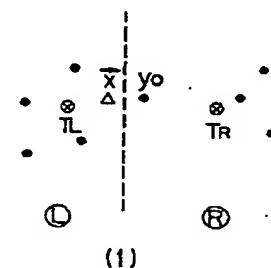
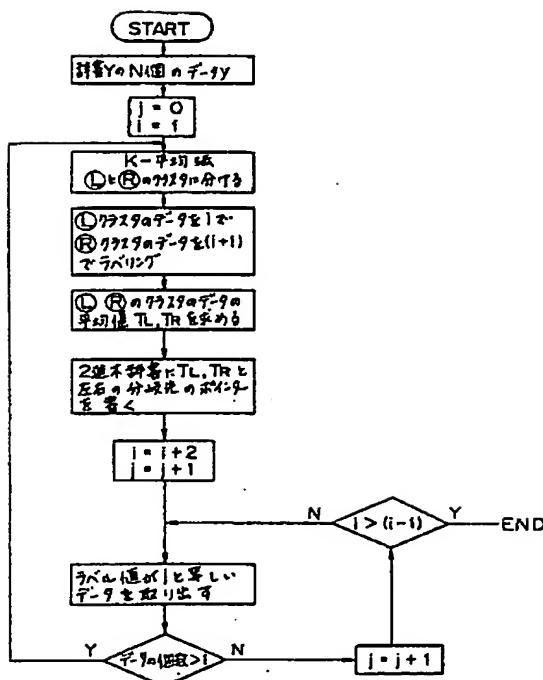
第1図



第 2 回



第3回



第 5 四